**A Project Report**

**On**

**Automate Diagnosis & Consultants (ADC)**

submitted for partial fulfillment of the requirements.

for the award of the degree of

**Bachelor of Technology**

in

**Computer Science**

**Submitted by**

Ankit Yadav (1900290120016)

Ayush Srivastava (1900290120029)

**Under the Supervision of**

Mr. Harsh Khatter

Assistant Professor  
Department of Computer Science



**KIET Group of Institutions, Ghaziabad**

**Dr. A.P.J. Abdul Kalam Technical University**

**2022-2023**

**DECLARATION**

We hereby declare that this submission is our work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material that to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature of Students

Name: Ankit Yadav

Roll No.: 1900290120016

Name: Ayush Srivastava

Roll No.: 1900290120029

Date: 06/04/2023

**CERTIFICATE**

This is to certify that Project Report entitled “**Automate Diagnosis & Consultant”** which is submitted by **Ankit Yadav, Ayush Srivastava & Deepanshu Singh** in partial fulfillment of the requirement for the award of degree B. Tech. in the Department of Computer Science of Dr A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

**Date:06/04/2023 Supervisor Signature**

Mr. Harsh Khatter

Assistant Professor  
Department of Computer Science

**ACKNOWLEDGEMENT**

It gives us great pleasure to present the synopsis of the B. Tech Mini Project undertaken during B.Tech. Fourth Year. We owe a special debt of gratitude to Mr. Harsh Khatter, Assistant Professor, Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness, and perseverance have been a constant source of inspiration for us. It is only his/her cognizant efforts that our endeavours have seen the light of the day.

We also take the opportunity to acknowledge the contribution of Dr. Ajay Kumar Srivastava, Head of the Department of Computer Science, KIET Group of Institutions, Delhi- NCR, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the department’s faculty members for their kind assistance and cooperation during the development of our project.

Last but not the least, we acknowledge our friends for their contribution to the completion of the project.

Signature:

Name: Ankit Yadav

Roll No: 1900290120016

Name: Ayush Srivastava

Roll No.:1900290120029

Date: 06/09/2022

**ABSTRACT**

A sizable segment of the world's population lacks access to quality healthcare. The success of healthcare ultimately depends on the doctor's skill. In this study, we investigate if this knowledge may be represented as an information corpus, or as data that has been retrieved using data mining methods, particularly the Machine Learning & Deep Learning Model, to make a diagnosis. When the medical diagnosis is made widely available, coverage increases and life quality improves. To determine whether inferences about the causes of various diseases can be made from the data, this paper provides an overview of machine learning approaches used in the classification of various diseases. We outline a few of our findings from the trials we ran before offering some suggestions for the future.

The difference between the current state of health and an acceptable or desirable health condition is the health problem. By lowering doctor visits, hospital stays, and diagnostic testing procedures, monitoring systems are designed to lower health care expenditures. Using the data mining modeling technique, the integration of clinical decision support with computer-based patient records could decrease medical errors, increase patient safety, stop unwelcome practice variance, and improve practice outcomes. Connecting patients and doctors through a user-friendly interface will make it easier for patients to use in emergency situations.

**TABLE OF CONTENTS**

|  |  |
| --- | --- |
|  | Page No. |
| TITLE PAGE .................................................................................................................... | 1 |
| DECLARATION .............................................................................................................. | 2 |
| CERTIFICATE …........................................................................................................... | 3 |
| ACKNOWLEDGEMENT.................................................................................................. | 4 |
| ABSTRACT...................................................................................................................... | 5 |
| LIST OF FIGURES ......................................................................................................... | 6 |
| TABLE OF CONTENTS………………………………………………………………... | 7 |
|  |  |
| CHAPTER 1 INTRODUCTION | 8 |
| 1.1.          Introduction ……………………................................................... |  |
| 1.2 Problem Statement.……………………....................................... |  |
| 1.2.          Objective………………………………………………………… |  |
| 1.3.          Scope……………………………………………………………. |  |
| CHAPTER 2 LITERATURE REVIEW…………………………………………….... | 9-11 |
| CHAPTER 3 PROPOSED METHODOLOGY …………………………………........ | 12-16 |
| 3.1 Flowchart |  |
| 3.2 Implementation Activity | 13-16  17 |
| CHAPTER 4 TECHNOLOGY USED ……….……………………….………………. | 18 |
| CHAPTER 5 DIAGRAMS …........................................................................................... |  |
| CHAPTER 6 CONCLUSION …....................................................................................... | 19 |
| REFERENCES….............................................................................................................. | 20-21 |

1. **INTRODUCTION**

**1.1 Introduction:**

The health problem is the gap between acceptable or desirable health status and the present status. In today’s life, health problems occurred more than last 25-30 years ago because of modernization, and industrialization. Suddenly, changes in the environment directly affect a health condition. So, more health-related problems are occurring day by day and require daily checkups on health conditions. This project is based on real-time implementation as well as more informative and realistic it can be highly used in Medical Diagnosis and understanding of the different feasibilities of the model.

The growth of ML has facilitated the increased use of computers in medicine. There are three primary goals of data mining that tend to be prediction, description, and presentation. Prediction involves some variables or fields in the dataset to predict unknown or future values of other variables of interest. The description focuses on finding patterns describing the data that can be interpreted by humans, on the other hand Presentation plays an important role to be easily understandable to humans.

Health issues are more prevalent today than they were 25 to 30 years ago due to urbanization and industrialization. Unexpected changes in the environment have a direct impact on health. As a result,  more health issues emerge daily, necessitating daily health condition checks. This project, which is based on real-time implementation and is more informative and realistic, can be very useful in helping to diagnose medical conditions.

The goal of creating automated systems is to decrease the amount of time and money spent on health care by lowering the number of doctor visits, hospital stays, and diagnostic testing procedures. The second main method is supervised learning, which uses machine learning techniques to learn a function from a collection of training data. Prediction and generalization are two key requirements for supervised learning algorithm performance. The trained function ought to be capable of accurately predicting the results for data that are not included in the training set. It should also serve as a model that generalizes to new data points and captures the underlying traits of the training data.

**1.2 Problem Statement:**

In such unprecedented times, with a fragile healthcare infrastructure, the medical facilities for people are scarce and even fewer for people with special needs. The healthcare sector got so preoccupied with Covid-19 cases that people with other ailments didn't even get a chance to avail medical assistance. So, we are here to present a model to mitigate the scarcity of facilities & the lack of medical management.

* Scarcity of facilities and the lack of medical management.
* There is no website for Automated Diseases Prediction System.
* Decentralized Patient Records are not available.
* No platform Integrates doctors and patients via disease prediction with relevant doctor suggestions.

**1.3 Field of Invention:**

The technology related to the general field of healthcare and management, i.e., machine learning-based automated medical diagnosis to enhance healthcare infrastructure. The difference between the current state of health and an acceptable or desirable health condition is the health problem. By lowering doctor visits, hospital stays, and diagnostic testing procedures, monitoring systems are designed to lower healthcare expenditures.

**1.4 Objectives:**

The objective of developing an automated system is to changing the way to presenting and detection of any diagnosis by

* Reduce health care costs by reducing physical hospital visits. hospitalizations, and diagnostic testing procedures.
* Integrate doctors with patients’ disease prediction with relevant doctor suggestions.
* Automated Diseases Prediction System to make facilities available easily.

**1.5 Scope:**

Setting up an alarm alert system & adding more diseases to our model increases our feasibility & compatibility with the market.

Adding more body parts models that increase people’s knowledge that would lead to training purposes for the nurses.

**1.6 Advantages:**

* Integration of clinical decision support with computer-based patient records could reduce medical errors and enhance patient safety.
* Unwanted practice variance is eliminated, and practice outcomes are improved, both of which can greatly raise the standard of clinical judgment.
* Reduce disease identification time and start treatment expeditiously.
* Reduce the cost of medical tests & by providing initial diagnostics in time that reduces huge amounts of the burden on the patient's family.
* Decentralized Patient Records lead to data being more accessible & makes the fast diagnosis of patients.
* Automated Diseases Prediction System leads to predicting disease on the basis of symptoms & recommending doctors & also book Path Labs on it.
* Integrate doctors and patients via an interactive and user-friendly interface so that Patients can easily utilize it in emergency situations.
* We also research pharmaceutical equipment & Medicines so that patients can get pure medicines from anywhere.

**CHAPTER 2 LITERATURE REVIEW**

**2.1 Machine learning techniques for classification of diabetes and cardiovascular diseases** Machine learning methods are playing an increasingly important role in data analysis because they can deal with massive amounts of data. In fact, the more data the better. Most machine learning methods construct hypotheses from data. our growing abilities to store large amounts of data in rapid-access computer memories and to compute with these data has enabled techniques that store and use all the data as they are needed.

These insights yield connections between deep learning and diverse physical and mathematical topics, including random landscapes, spin glasses, jamming, dynamical phase transitions, chaos, Riemannian geometry, random matrix theory, free probability, and nonequilibrium statistical mechanics. Indeed, the fields of statistical mechanics and machine learning have long enjoyed a rich history of strongly coupled interactions, and recent advances at the intersection of statistical mechanics and deep learning suggest these interactions will only deepen going forward.

**2.2 Supervised learning & Unsupervised Learning**

ML methods that aim to learn a function from a given training data set constitute the second main approach, called supervised learning. Two main criteria for the success of supervised learning algorithms are prediction and generalization. The learned function should be able to successfully predict the output for data other than in the training set used. At the same time, it should act as a model that captures the underlying characteristics of the training data and generalize to new data points.

Supervised ML techniques offer an invaluable set of formalized computing methods to develop computer-assisted detection, analysis, and decision systems for network security. These methods have been successfully applied to a wide variety of fields ranging from image recognition, speech processing, and data mining. the opposite behaviour is called overfitting where the function describes the training data perfectly but has poor predictive power. The performance of a supervised learning algorithm is quantitatively assessed using a test data set and cross-validation techniques.

**2.3 Regression Techniques**

Linear regression analysis is often used by life scientists. For example, the equation for the regression of one variable on another may suggest hypotheses about why the two variables are functionally related. More practically, regression can be used in situations where the dependent variable is difficult, expensive, or impossible to measure, but its values can be predicted from another easily measured variable to which it is functionally related. Logistic regression is an extension of linear regression. Rather than modelling a linear relationship between the independent variable x and the probability of the outcome, which is unnatural since it would allow predicted probabilities outside the range of 0–1, it assumes a linear (straight line).

A major advantage of logistic regression compared to other similar approaches like probity regression—and therefore, a reason for its popularity among medical researchers—is that the exponentiated logistic regression slope coefficient (*e*b) can be conveniently interpreted as an odds ratio. The odds ratio indicates how much the odds of a particular outcome change for a 1-unit increase in the independent variable (for continuous independent variables) or versus a reference category.

**2.4 Support Vector Machine in medical fields**

Support Vector Machine is a supervised Machine Learning Algorithms that can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is several features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.

The SVM kernel is a function that takes low dimensional input space and transforms it to a higher dimensional space i.e., it converts not separable problem to separable problem. It is mostly useful in non-linear separation problem. Simply put, it does some extremely complex data transformations, then finds out the process to separate the data based on the labels or outputs you’ve defined.

**2.5 Research and Analysis of the Front-end Frameworks and Libraries in E-Business Development**

With web technology rapidly expands out in recent years, there is a significant trend that Hypertext Markup Language (HTML)5 turns into a worldwide web consortium and leads the front-end development to stand on the front stage of internet history. However, there are numerous front-end development frameworks and libraries such as React, Angular and Vue. How to select a suitable framework or library to establish the e-Business and reach out to maximize the user experience becomes a priority operation in web development. This paper starts with introducing an overview of the leading frameworks and libraries in the field of front-end development and examine each performance in web services.

**2.6 A review and analysis of technologies for developing web Applications**

Web applications tend to be multi-tiered by nature, with the most common structure being the three- tiered architecture. In its most common form, the three tiers are • Presentation layer: The presentation tier is the front-end layer in the 3- tier system and consists of the user interface. This user interface is often a graphical one accessible through a web browser or web-based application and which displays content and information useful to an end user. The four big technologies in this layer are HTML, CSS, JavaScript and jQuery. The workflows by which the data and requests travel through the back end are encoded in a business layer. Scripting languages that are uses are ColdFusion, Ruby, WebObjects, and Python.

**2.7 Research on HTML5 in Web Development**

The purpose of this study was about HTML (Hypertext Markup Language) is the code that is used to structure a web page and its content. It is the building block of a website. HTML5 is the next major revision of the HTML standard superseding HTML 4.01, XHTML 1.0, and XHTML HTML5 is a standard for structuring and presenting content on the World Wide Web. HTML5 is a cooperation between the World Wide Web Consortium (W3C) and the Web Hypertext Application Technology Working Group (WHATWG). The lesson was about new tags that are introduced in HTML5. With the development of HTML5 it has wide range of applications in multimedia direction. It can play audio and video and supports animations from the browser without the need of the proprietary technologies. HTML5 introduces a number of new elements and attributes that can help you in building modern websites.

**2.8 Challenges in Android Application Development**

The following are the challenges faced by the Android App developers The most common challenge is to set the properties of app for different devices with different screen sizes, resolution etc. There are many versions of each android device and while releasing the app, version specific details need to be checked this makes the task very critical. Currently, the Development Environment does not have enough tools for testing. There is a need of testing techniques for the Android Platform. Also debugging features must be made available. Sometimes different devices have different capabilities in terms of software support like some browsers has poor support for HTML5.

**CHAPTER 3 UML DIAGRAMS**

Various embodiments of the invention are disclosed in the following detailed description & accompanying drawings:

Diagram, schematic

Description automatically generated

Fig 1: Component Diagram

Diagram

Description automatically generated

Fig 2.1: Activity Diagram (PathLabs module).

Diagram

Description automatically generated

Fig 2.2: Activity Diagram (Doctors module).

Diagram

Description automatically generated

Fig 2.3: Activity Diagram (Patients module).

A picture containing diagram

Description automatically generated

Fig 3.1: Data Flow Diagram (Level -0)

Diagram

Description automatically generated

Fig 3.2: Data Flow Diagram (Level -1 Patient Module)

Diagram

Description automatically generated

Fig 4: Entity Relationship Diagram

Diagram

Description automatically generated

Fig 5.1: Login Module

**Diagram

Description automatically generated**Fig 5.2 : Forgot Password Module

**Diagram

Description automatically generated**

Fig 6.1: Use Case Diagram of doctor & patient Module

Diagram

Description automatically generated

Fig 6.2: Use Case Diagram of Path labs & Admin Module

A picture containing graphical user interface

Description automatically generated

Fig 7: Internet of Things Module of Automated Diagnosis.

**Shape

Description automatically generated**

Fig 8: Data Streaming Model

**A picture containing diagram

Description automatically generated**

Fig 9: Use case model of the Frontend link with ML model.

**CHAPTER 4 METHODOLOGY**

The methodology's goal is to forecast an individual's risk of developing kidney, lung, and breast cancer, heart disease, and diabetes using a few questions and machine learning models in an end-to-end procedure. the system I used for my research has the following software and system configurations: On an Intel(R) Core(TM) i5-2310M GPU @1650Ti with 8 GB RAM, Jupyter Notebook 5.5.0 and VS Code 1.73 are used to implement Python 3 and the Flask framework.

Figure 1 displays a block schematic of the fundamental procedures used for each machine-learning model. To transform the raw data into a form that can be used, data cleaning is done first. Data analysis is carried out after data cleansing to ascertain the significance of characteristics. Once a goal has been established, it is time to start gathering the data required for analysis. Your data team will be responsible with cleaning and sifting through the data once it has been gathered from all the required sources. Because not all data is good data, data cleansing is crucial during the data analysis process. Data mining, which is referred to as "knowledge discovery inside databases," is one method. In order to predict what will probably happen next in the future, predictive analysis looks ahead to the future. These methods are a component of inferential statistics, which is the act of examining statistical data in order to make inferences about the connections between various sets of data. We have to utilize the Train/Test methodology to evaluate the model's performance. The train/test approach is a way to gauge how accurate your model is. Because you divide the data set into two sets, a training set and a testing set, the method is known as Train/Test. 20% for testing, 80% for training. Using the training set, you train the model. Utilizing the testing set, you test the model. Create the model by training it. To test a model is to determine its correctness. After confirming that our model is sound, we can now begin making fresh value predictions.

Integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decease unwanted practice variation and improve practice outcome which can help significantly improve the quality of clinical decisions using the data mining modelling technique & Integrate doctors and patients via interacting and user-friendly interface so that Patient can easily utilize in emergency situations. There are three primary goals of this model tend to be prediction, description, and presentation.

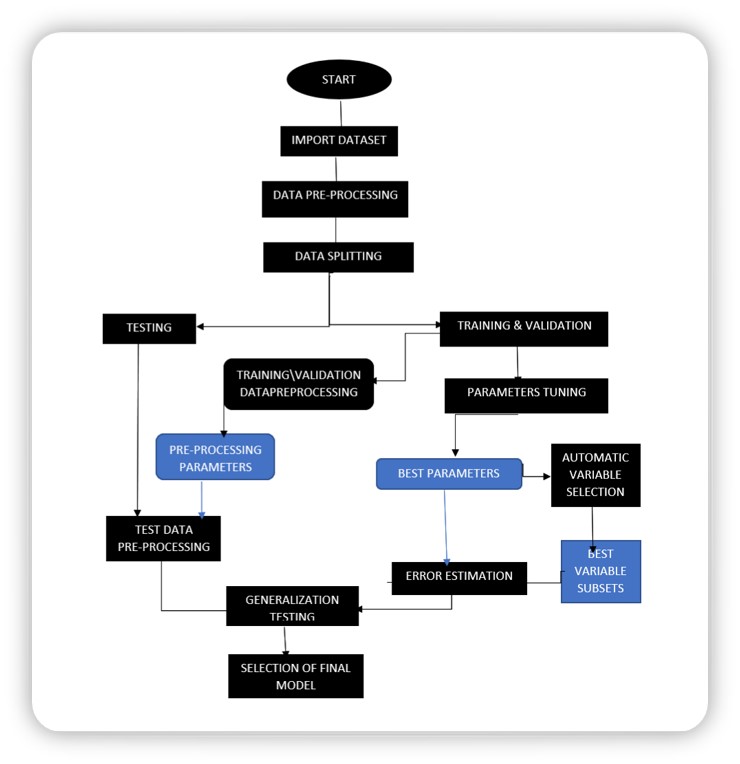


Fig 10: Machine Learning Algorithm’s flow chart

**4.1 Algorithm for the proposed system:**

**Random Forest Classifier Algorithm:**   
A group of several decision trees is called a random forest. Decision trees are employed as parallel estimators in the bagging technique, which is used to construct random forests. When used in a classification issue, the outcome is determined by the majority vote of the findings from each decision tree. In a regression, the mean value of something like the target values in a leaf node serves as the prediction. The mean value of the decision tree outcomes is taken into account by random forest regression.

**Logistic Regression:**A set of independent variables is used to estimate discrete values (often binary values like 0/1) using logistic regression. By fitting data to a logit function, it aids in the prediction of an event's likelihood. Logistic regression is another name for it. The following techniques are frequently used to enhance logistic regression models: incorporate interaction terms, remove features, regularize approaches, and employ a non-linear mode.

**Decision Tree:**  
The Decision Tree method, a supervised learning technique used for issue classification, is one of the most widely used algorithms in machine learning today. Both continuous and categorical dependent variables may be classified using it. Based on the most important characteristics/independent variables, this method splits the population into two or more homogenous groupings.

**Step 1**: Gathering and identifying the data that will be provided to the network as input is the first stage in the categorization of diabetes or heart disease using Logistic Regression, Decision Tree & Random Forest.   
**Step 2**: The network receives a defined training dataset and the selected training method. After the training phase, the Logistic Regression, Decision Tree  & Random Forest are also put to the test to get feedback on how well they categorize the condition.

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Algorithm** | **Function** | **Use** |
| Basic Regression | Logistic Regression | linear\_model.LogisticRegression() | The target variable is categorical. |
| Classification | Decision Tree | tree.DecisionTreeClassifier() | If/then/else. Non-contiguous data. can also be regression. |
| Classification | Random Forest | ensemble.RandomForestClassifier() | Find the best split randomly. can also be regression. |

**Step 3:** The "training data set" is used to train the classifier, the "validation set" is used to fine-tune the parameters, and the "test data set" is used to evaluate the performance of the classifier. It's vital to keep in mind that only the training and/or validation set is available when the classifier is being trained. The test data set must not be used for classifier training. Only when the classifier is being tested will the test set be accessible.

**Step 4:** We may then create a confusion matrix, which indicates how effectively our model has been trained after this is complete. True Positives, True Negatives, False Positives, and False Negatives are the four parameters that make up a confusion matrix. In order to create a model that is more accurate, we would want to obtain more data for the true negatives and true positives. The number of classes has a direct impact on the size of the confusion matrix.  
**Step 5:** The model creation process includes a step called model evaluation. Finding the model that best depicts our data and predicts how well the model will perform in the future is helpful. In order to enhance the model, we may adjust its hyper-parameters to attempt to raise its accuracy while also looking at the confusion matrix to try to increase the proportion of true positives and true negatives.

Diagram

Description automatically generated

Fig 11: System flow Diagram

**4.2 Proposed System of Automate Diagnosis & Consultant:**

The dashboard will run the function that connects patients and doctors. The dashboard has been designed to get a better understanding of a patient's condition. The medical profession is being transformed by technological advancements, which is causing a rapid change in the medical business. Artificial intelligence (AI) is being used to diagnose illnesses, predict diseases, and develop treatments. In this article, it will be discussed how AI can support the management of level-three 911 calls and assist physicians in making better patient decisions. According to the doctors' available time slots, those who have appointments can show up for their appointments. To avoid any unneeded waiting times, this is done. Modern medical professionals are employing cutting-edge technology to provide more precise diagnoses and individualized treatment strategies. Doctors can now give patients extensive medical data, an in-depth report on their symptoms, and personalized care plans that can be seen on a mobile device thanks to artificial intelligence (AI). With the use of an app, the doctor may view the patient's information, write him a prescription, and suggest testing. This software assists in quickly diagnosing the patient since it has all the information required to understand the ailment, its symptoms, and recommended treatments. The outdated paper lab slip is being replaced with a modern dashboard. Patients will be able to schedule an appointment at a time that works for them because these slips are now digital. As soon as the reports are finished being processed by the laboratory, they will also be automatically uploaded into the patient's database. Doctors use information from a variety of sources to provide the most precise diagnosis possible. With the more recent use of electronic medical records, a more complete image of patients is now possible, improving care.

* **Scalability:**
  + The primary objective of a SWOT analysis is to help organizations develop a full awareness of all the factors involved in making a business decision.
  + Strength: Small (4-5 employees) can change and adapt quickly & gives a unique advantage.
  + Weakness: Team members who are not much trained in both technical and professional engineering. No one has been through any such type of technical training programs.
  + Opportunity: Market trends & Economic trends will take us to new heights.
  + Threats: Political, environmental, economic regulations & Funding part is our main threat.
* **How Realistic to Achieve/Make?**
  + It can be easily implementable because in this pandemic period most of peoples are technically advanced & they are mostly relying on the internet.
  + It can be widely spread on different social media platforms, Application based platforms like google play store, Appstore etc.
  + It can be easily accessed & modified.
* **Any Threat/Risk/Problem that you can foresee?**

Bad Design, Poor user Experience, Slopy implementation, feature creep, & lack of quality control all contribute to product failure.

**4.3 Implementation activity:**

Disease diagnosis using Machine Learning & Deep Learning Models

• This project focused on many health issues from which not only India but the whole World is suffering. Visualize the dataset & using All ML & DL algorithms to diagnose Diseases.

• Dataset used: Kaggle Cardiovascular, diabetes, Kidney, Breast Cancer, malaria etc. dataset

• Project GitHub Link: https://github.com/ankiii07/Automate-Diagnosis

* **Economic Sustainability:**
  + We first go personally to many hospitals for our product launching & later with some remote areas so that they use our application & get more & more benefits from it from anywhere.
  + At Initial Stage we just put more & more efforts to setup in market not on generating revenue.
  + it’s not for generating revenue but later after this product get popular and provide more & more benefits to people, then we take charges for using this application which is affordable to our middle order people.
* **Environment Sustainability:**
  + The idea is much innovative and not implemented with such mixed technology so for which made the people to their comfort and so guide them to reach their desire.
  + Apart from this it is an initiative which has not been seen so far and it has Less development and maintenance cost and high profit values.
  + This Project is Environment Sustainable & much helpful to Remote areas public, rural area public & it does not cause any environmental damage.

**CHAPTER 5 TECHNOLOGY USED**

* **Web Development**

**Frontend**: HTML, CSS, JavaScript.  
**Backend**: NodeJs, ExpressJs MongoDB.

* **Machine Learning & Deep Learning**

Anaconda, Jupyter Notebook, EDA, Data Collection, Cleaning, Visualisation, NumPy, Pandas, Seaborn, SciPy, Matplotlib, Pandas, Sklearn Libraries, ML Algorithms

* **Augmented Reality**

Unity, Vuforia SDK, AR Core, A Frame, JDK, Augmented Reality integrated with Machine Learning

* **Internet of Things**

Raspberry Pi4, GSM module, Sensors

* **Blockchain**

.

**CHAPTER 6 SUMMARY**

ADC brings you an online platform, which can be accessed for all your health needs. We are trying to make healthcare a hassle-free experience for you.

Lab tests? That too in the comfort of your home.

Doctor consult? Yes, we got that covered too.

In healthcare, the information about our medicines and lab tests is either unavailable or incomprehensible to us.

We provide accurate, authoritative & trustworthy information on medicines and help people use their medicines effectively and safely.

We also provide diagnostic services from certified labs and online doctor consults at any time, from anywhere.

Till now, we decided on the Title of the project, End Users, and did an analysis of the existing software whether it is in India or Outside India. features of the product that we give to the end users, try to make the maximum of these.

We move on to the use case of the product regarding the Patient, Doctors, Labs & their Architecture. Data flow Diagram of the product with their use cases & talk about the database, then moving to the Data Flow Diagram level-2 & we have to make Data Dictionary & Data Schema of the Doctor, Patients, & Labs as well as Admin of the product that approved everything.

We make the website using Reacts & Application with the help of Flutter compatible with Both Android as well as IOS Users. Designing Screens of the application has been done so far.

**CHAPTER 7 CONCLUSION**

As healthcare services are an important part of our society, automating these services lessen the burden on humans and diseases through the measuring process. Also, the transparency of this system helps patients to trust it. When the threshold value is reached, the alarm system that consists of a buzzer and LED alerts the doctor and they can act more quickly.

The IoT technology helps the server to update the patient data on the website. Many further improvements can be made to our system to make it better and easily adaptable such as adding more advanced sensors. The biometric information of the patient which is stored and published online & transmit to cellular can be given to scientists and researchers in medical fields to analyze the value and find patterns or for other research work.

Future work will focus on monitoring additional health-related parameters using a broader combination of transducers, sensors, and correlation techniques, and on improving system reliability and robustness to patient movement and connectivity losses.

**CHAPTER 8 RESEARCH PAPER**

**Machine learning-based automated medical diagnosis to enhance healthcare infrastructure.**

**ABSTRACT**

A sizable segment of the world's population lacks access to quality healthcare. The success of healthcare ultimately depends on the doctor's skill. In this study, we investigate if this knowledge may be represented as an information corpus, or as data that has been retrieved using data mining methods, particularly the Machine Learning & Deep Learning Model, to make a diagnosis. When the medical diagnosis is made widely available, coverage increases and life quality improves. In order to determine whether inferences about the causes of various diseases can be made from the data, this paper provides an overview of machine learning approaches used in the classification of various diseases. We outline a few of our findings from the trials we ran before offering some suggestions for the future.

The difference between the current state of health and an acceptable or desirable health condition is the health problem. By lowering doctor visits, hospital stays, and diagnostic testing procedures, monitoring systems are designed to lower health care expenditures. Using the data mining modeling technique, the integration of clinical decision support with computer-based patient records could decrease medical errors, increase patient safety, stop unwelcome practice variance, and improve practice outcomes. Connecting patients and doctors through a user-friendly interface will make it easier for patients to use in emergency situations.

**INTRODUCTION**

Machine learning (ML) is a branch of artificial intelligence that uses "learning ability provided to computers without additional programming" to address problems in the real world. Research into whether computers could learn to emulate the human brain led to the development of machine learning. When Arthur Samuel created the first checkers game-playing software in 1952, ML made its initial attempts to generate the necessary skills to defeat a world champion. Later in 1957, Frank Rosenblatt developed an electronic system that can mimic the way the human brain works to learn how to handle complex issues.

The growth of ML has facilitated the increased use of computers in medicine.There are three primary goals of data mining that tend to be prediction, description, and presentation. Prediction involves some variables or fields in the dataset to predict unknown or future values of other variables of interest. Description focuses on finding patterns describing the data that can be interpreted by humans, on the other hand Presentation plays an important role to be easily understandable to humans.

Health issues are more prevalent today than they were 25 to 30 years ago due to urbanization and industrialization. Unexpected changes in the environment have a direct impact on health. As a result, more health issues emerge daily, necessitating daily health condition checks. This project, which is based on real-time implementation and is more informative and realistic, can be very useful in helping to diagnose medical conditions and comprehend the many model feasibility options.

The availability of medical facilities for people is limited under these unheard-of economic conditions, and it is even more so for those with specific requirements. People with other illnesses didn't even get an opportunity to access medical aid because the healthcare industry became so preoccupied with Covid-19 instances. We are here to provide a model that will help alleviate the lack of facilities and  poor medical management.

The goal of creating automated systems is to decrease the amount of time and money spent on health care by lowering the number of doctor visits, hospital stays, and diagnostic testing procedures. The second main method is supervised learning, which uses machine learning techniques to learn a function from a collection of training data. Prediction and generalization are two key requirements for supervised learning algorithm performance. The trained function ought to be capable of accurately predicting the results for data that are not included in the training set. It should also serve as a model that generalizes to new data points and captures the underlying traits of the training data.

Machine learning is frequently used to categorize diseases, and scientists are becoming increasingly more interested in creating such systems for many disease diagnosis and tracking as well. Diabetes, cardiovascular disease & many more are the top 10 killers worldwide, according to the World Health Organization (WHO). According to a study from January 2017, CDs are the leading cause of death worldwide. In the list of the top 10 causes of death over the past 15 years, the world's worst disease, which claimed 15 million lives in 2015, holds the top spot.

**LITERATURE SURVEY:**

Biosignals from patients have been used in the past to construct AI systems for clinical decision assistance. Such organized clinical data includes unprocessed signals.  a lack of sufficient background for appropriate interpretation, whereas clinical publications with unstructured free text contain comprehensive explanations of broader clinical situations.  
   
Tamilselvan. P [1]. Monitors based on blood pressure and ECG readings are available. Reactions are kept the signals that these sensors transmit to the Using a signal conditioner and amplifier, Raspberry Pi because the signals' levels are low (gain), an amplifier is necessary (scu). The signals that these sensors transmit to the Raspberry Pi are processed by a signal conditioner and amplifier, which is required because the signal levels are low (gain). Using a circuit, the signal is amplified and sent to a Raspberry Pi. Linux is used to run the Raspberry Pi computer. The system operates like a minicomputer processing system. Here, patients' ECG and blood pressure are measured with the appropriate sensors. Moreover, it is monitorable on a computer's monitor. Utilizing a Raspberry Pi and monitoring from anywhere, Internet sources are used everywhere.

Vivek Datla, Sadid A. Hasan [2]. outlines our Knowledge Graph (KG)-system for based clinical diagnostic inference. We performed substantial testing on the MIMIC-III benchmark dataset, analyzing different parts of a clinical note. Results proved that the details of the current illness's history were relevant. The parts on prior health histories often offer the greatest insight. inference of a clinical diagnosis in comparison to all portions. Furthermore, we demonstrated that the KG-based system can perform admirably with a loose accuracy metric in comparison to the cutting-edge CMemNN model.

Hiroshi Sugimura, Kazuki Utsumi [3]. They suggest a method that superimposes online service information onto the daily environment. Natural behaviors like speech and gestures are used to operate the system. The system is put together using three input/output devices: a microphone for sound recognition, a camera for gesture detection, and a projector for information presentation. We provided a detailed account of the conception, design, execution, and assessment of a prototype system. The proposal system's value was then confirmed.

Fabio Santos, Filipe Silva and Petia Georgieva [4]. focused on the skin lesion diagnosis techniques integrated inside eHealth apps that help individuals and medical professionals and are clearly needed as the prevalence of skin cancer grows. Meanwhile, recent developments in deep learning techniques enable performance that is close to that of a dermatologist and has a large room for growth, outperforming previous approaches. Before putting such tools into use in the real world, issues like the need for large datasets or the high computing needs must be resolved because they negatively affect how well models function. However, effective methods like these reduce their impacts through transfer learning and data augmentation, according to research. Finally, it is anticipated that when more information on skin lesions is made publicly available, these difficulties will lose some of their significance.

Berina Ali [5] provides an overview of machine learning methods for categorizing CVD and diabetes using artificial neural networks (ANNs) and Bayesian networks (BNs). A comparative study was carried out on a few publications released throughout the time period, between 2008 and 2017. In a few articles, multilayer feedforward neural networks using the Levenberg-Marquardt learning method are the most commonly utilized ANN type. Additionally, utilizing ANN improved the computation of the mean accuracy of observed networks, indicating a greater likelihood of obtaining more accurate findings for the categorization of CVD and/or diabetes.

**METHODOLOGY:**

The methodology's goal is to forecast an individual's risk of developing kidney, lung, and breast cancer, heart disease, and diabetes using a few questions and machine learning models in an end-to-end procedure. the system I used for my research has the following software and system configurations: On an Intel(R) Core (TM) i5-2310M GPU @1650Ti with 8 GB RAM, Jupyter Notebook 5.5.0 and VS Code 1.73 are used to implement Python 3 and the Flask framework.

Figure 1 displays a block schematic of the fundamental procedures used for each machine-learning model. To transform the raw data into a form that can be used, data cleaning is done first. Data analysis is carried out after data cleansing to ascertain the significance of characteristics. Once a goal has been established, it is time to start gathering the data required for analysis. Your data team will be responsible with cleaning and sifting through the data once it has been gathered from all the required sources. Because not all data is good data, data cleansing is crucial during the data analysis process. Data mining, which is referred to as "knowledge discovery inside databases," is one method. In order to predict what will probably happen next in the future, predictive analysis looks ahead to the future. These methods are a component of inferential statistics, which is the act of examining statistical data in order to make inferences about the connections between various sets of data. We have to utilize the Train/Test methodology to evaluate the model's performance. The train/test approach is a way to gauge how accurate your model is. Because you divide the data set into two sets, a training set and a testing set, the method is known as Train/Test. 20% for testing, 80% for training. Using the training set, you train the model. Utilizing the testing set, you test the model. Create the model by training it. To test a model is to determine its correctness. After confirming that our model is sound, we can now begin making fresh value predictions.

**ALGORITHM OF THE PROPOSED SYSTEM:**

**Random Forest Classifier Algorithm:**   
A group of several decision trees is called a random forest. Decision trees are employed as parallel estimators in the bagging technique, which is used to construct random forests. When used in a classification issue, the outcome is determined by the majority vote of the findings from each decision tree. In a regression, the mean value of something like the target values in a leaf node serves as the prediction. The mean value of the decision tree outcomes is taken into account by random forest regression.

**Logistic Regression:**A set of independent variables is used to estimate discrete values (often binary values like 0/1) using logistic regression. By fitting data to a logit function, it aids in the prediction of an event's likelihood. Logistic regression is another name for it. The following techniques are frequently used to enhance logistic regression models: incorporate interaction terms, remove features, regularize approaches, and employ a non-linear mode.

**Decision Tree:**  
The Decision Tree method, a supervised learning technique used for issue classification, is one of the most widely used algorithms in machine learning today. Both continuous and categorical dependent variables may be classified using it. Based on the most important characteristics/independent variables, this method splits the population into two or more homogenous groupings.

**Step 1**: Gathering and identifying the data that will be provided to the network as input is the first stage in the categorization of diabetes or heart disease using Logistic Regression, Decision Tree & Random Forest.  
**Step 2**: The network receives a defined training dataset and the selected training method. After the training phase, the Logistic Regression, Decision Tree & Random Forest are also put to the test to get feedback on how well they categorize the condition.  
**Step 3:** The "training data set" is used to train the classifier, the "validation set" is used to fine-tune the parameters, and the "test data set" is used to evaluate the performance of the classifier. It's vital to keep in mind that only the training and/or validation set is available when the classifier is being trained. The test data set must not be used for classifier training. Only when the classifier is being tested will the test set be accessible.  
**Step 4:** We may then create a confusion matrix, which indicates how effectively our model has been trained, after this is complete. True Positives, True Negatives, False Positives, and False Negatives are the four parameters that make up a confusion matrix. In order to create a model that is more accurate, we would want to obtain more data for the true negatives and true positives. The number of classes has a direct impact on the size of the confusion matrix.  
**Step 5:** The model creation process includes a step called model evaluation. Finding the model that best depicts our data and predicts how well the model will perform in the future is helpful. In order to enhance the model, we may adjust its hyper-parameters to attempt to raise its accuracy while also looking at the confusion matrix to try to increase the proportion of true positives and true negatives.

**Proposed System:**

The dashboard will run the function that connects patients and doctors. The dashboard has been designed to get a better understanding of a patient's condition. The medical profession is being transformed by technological advancements, which is causing a rapid change in the medical business. Artificial intelligence (AI) is being used to diagnose illnesses, predict diseases, and develop treatments. In this article, it will be discussed how AI can support the management of level-three 911 calls and assist physicians in making better patient decisions. According to the doctors' available time slots, those who have appointments can show up for their appointments. To avoid any unneeded waiting times, this is done. Modern medical professionals are employing cutting-edge technology to provide more precise diagnoses and individualized treatment strategies. Doctors can now give patients extensive medical data, an in-depth report on their symptoms, and personalized care plans that can be seen on a mobile device thanks to artificial intelligence (AI). With the use of an app, the doctor may view the patient's information, write him a prescription, and suggest testing. This software assists in quickly diagnosing the patient since it has all the information required to understand the ailment, its symptoms, and recommended treatments. The outdated paper lab slip is being replaced with a modern dashboard. Patients will be able to schedule an appointment at a time that works for them because these slips are now digital. As soon as the reports are finished being processed by the laboratory, they will also be automatically uploaded into the patient's database. Doctors use information from a variety of sources to provide the most precise diagnosis possible. With the more recent use of electronic medical records, a more complete image of patients is now possible, improving care.

**DISCUSSION:**  
Data mining techniques play a significant role in medical systems, which will significantly contribute to the advancement of the medical industry. This paper provides a disease categorization based on several data mining and artificial intelligence technologies. Furthermore, we discovered in the literature that there are two kinds of factors utilized in disease categorization.

We noticed in the literature that the  classification of illnesses still has room for improvement. Despite the fact that proper characteristics may be derived from the ECG, it is a noninvasive strategy used to diagnose patients, and the ECG signal does not provide the necessary information. Because biosignals have an irregular structure, creating an effective technique for hidden factor extraction from ECG signals is particularly challenging. Several studies have found that the feature extraction method is incapable of determining the exact values of unmasked ECG signal parameters. Furthermore, using a restricted dataset for classification may result in misclassification; hence, in order to overcome the error rate, it is critical to avoid using a short dataset for classification.

**CONCLUSION:**

Because healthcare services are such an important aspect of our society, automating them relieves human stress while also making measurement simpler. Furthermore, the system's transparency promotes patient trust. When the threshold value is achieved, an alarm system consisting of a buzzer and an LED alerts the physicians, allowing them to respond more quickly.

This work focused on a variety of health concerns that affect not only India but the entire world. Visualize the dataset and identify diseases using all ML and DL algorithms. Integrate physicians and patients with an interactive and user-friendly interface so that patients may use it in an emergency. This model's three key purposes are prediction, description, and presentation. The patient's biometric data, which is captured, published online, and transferred to cellular devices, may be made available to scientists and researchers in medical disciplines in order to evaluate its value and reveal patterns, as well as for other research objectives.

Future research will focus on monitoring additional health-related indicators with a bigger collection of transducers, sensors, and correlation algorithms, as well as improving system dependability and resilience in the face of patient movement and connectivity losses.

**REFERENCES**

[1] D. Fagella, 7 Applications of Machine Learning in Pharma and Medicine, 2017, [online]

[2] V. S. Pendyala and S. Figueira, "Automated Medical Diagnosis from Clinical Data," 2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService), 2017, pp. 185-190, doi: 10.1109/BigDataService.2017.14.

[3] F. Santos, F. Silva and P. Georgieva, "Automated Diagnosis of Skin Lesions," 2020 IEEE 10th International Conference on Intelligent Systems (IS), 2020, pp. 545-550, doi: 10.1109/IS48319.2020.9200090.

[4] V. Datla et al., "Automated clinical diagnosis: The role of content in various sections of a clinical document," 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2017, pp. 1004-1011, doi: 10.1109/BIBM.2017.8217794.

[5] J. Siddiqi, B. Akhgar, A. Gruzdz, G. Zaefarian and A. Ihnatowicz, "Automated Diagnosis System to Support Colon Cancer Treatment: MATCH," Fifth International Conference on Information Technology: New Generations (itng 2008), 2008, pp. 201-205, doi: 10.1109/ITNG.2008.62.

[6] P. Ongsulee, "Artificial intelligence, machine learning and deep learning," 2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE), 2017, pp. 1-6, doi: 10.1109/ICTKE.2017.8259629.

[7] R. Chellappa, S. Theodoridis and A. van Schaik, "Advances in Machine Learning and Deep Neural Networks," in Proceedings of the IEEE, vol. 109, no. 5, pp. 607-611, May 2021, doi: 10.1109/JPROC.2021.3072172.

[8] F. Ertam and G. Aydın, "Data classification with deep learning using Tensorflow," 2017 International Conference on Computer Science and Engineering (UBMK), 2017, pp. 755-758, doi: 10.1109/UBMK.2017.8093521.

[9] H. Božiković and M. Štula, "Web design — Past, present and future," 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018, pp. 1476-1481, doi: 10.23919/MIPRO.2018.8400266.

[10] D. S. Blyth, "Web page design for HTML (and friends)," IPCC 96: Communication on the Fast Track. IPCC 96 Proceedings, 1996, pp. 89-103, doi: 10.1109/IPCC.1996.552585.

[11] H. Ran, W. Zhuo and X. Jianfeng, "Web Quality of Agile Web Development," 2009 IITA International Conference on Services Science, Management and Engineering, 2009, pp. 426-429, doi: 10.1109/SSME.2009.112.

[12] S. Drobi, "Play2: A New Era of Web Application Development," in IEEE Internet Computing, vol. 16, no. 4, pp. 89-94, July-Aug. 2012, doi: 10.1109/MIC.2012.84.

[13] H. Sugimura et al., "Development of immersive display system of web service in living space," 2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE), 2014, pp. 49-50, doi: 10.1109/GCCE.2014.7031106.

[14] N. Sandhya and K.R. Charanjeet, "A review on Machine Learning Techniques", International Journal on Recent and Innovation Trends in Computing and Communication, pp. 395-399, 2016, ISSN 2321-8169.